Comparison between Japanese and English Tweets about Greta Thunberg

Context

  Frequently I use twitter, and after the news that Greta Thunberg addressed environmental problems, I noticed that many people mentioned Greta Thunberg with their negative view in contrast that the number of people who mentioned with their positive standpoint was few.

 However, some people said that the information from out of Japan was totally opposite to Japan. They said most people regarded the address by Greta as a positive thing and movements for environmental problems by young people become more and more popular.

 I'm Japanese. Hence my thought, such as my view isn't likely to be neutral, is natural. Therefore, I'm interested in this comparison.

Research question

Try to identify whether existing the differences about responses to Greta or not.

Data and method

I used 7999 English and 9991 Japanese tweets.

To get the data, I searched tweets that contains "Greta" and "グレタ". "グレタ" is equal to Greta. In this time, I didn't use hashtags because there is a tendency that Japanese users retweet a Tweet which contains a topic of which they want to talk about when they are talking about it.

The data was fetched at 2019/10/13 around pm 20:10.

I analyzed them with sentimental analysis.

Each word has a value between -1 with negative to +1 with positive meanings, so, I classified them with positive word which has more than 0 and negative word which has less than 0 for Japanese and English.

To split each word, I used RMeCab. We can split Japanese words by one morpheme.

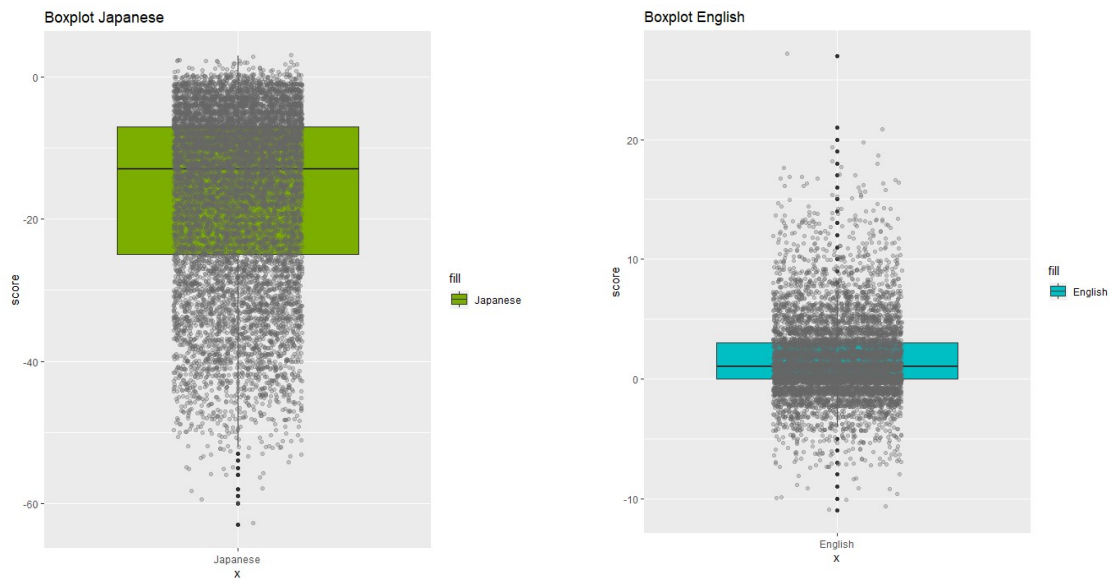Page: http://www.lr.pi.titech.ac.jp/~takamura/pndic_en.html
Japanese Dictionary: http://www.lr.pi.titech.ac.jp/~takamura/pubs/pn_ja.dic
English Dictionary: http://www.lr.pi.titech.ac.jp/~takamura/pndic_en.html
positive_negative_splitter: https://github.com/tori-takashi/positive_negative_splitter

As for Japanese tweets, I observe that the bot tweets a lot about Greta speech, therefore I rejected tweets from twittbot.net in observations. The total tweet in Japan is 9804.

Results

As we can see, Japanese tweets and English tweets have totally different distributions.

The mean value of Japanese was -16.715 and English one was +1.57.

Therefore, we can conclude that the response for Greta speech was totally different between Japanese and using English users.

Discussion

Although this comparison conclude that the distribution is totally different, we have a problem about the way to compare.

The problem is the weight for each word.
I separated between positive and negative but the any weight didn't consider. It means that I have regarded as a same weight for the value 0.02 and 1.0.

We have a problem about the way to compare though, we can say that Japanese might be tend to regard Greta as negative compared with global users.